

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications, Department of Child,
Youth, and Family Studies

Child, Youth, and Family Studies, Department of

2013

An Introduction to Latent Variable Mixture Modeling (Part 2): Longitudinal Latent Class Growth Analysis and Growth Mixture Models

Kristoffer S. Berlin

University of Memphis, kristoffer.berlin@gmail.com

Gilbert R. Parra

University of Southern Mississippi, gparra2@unl.edu

Natalie A. Williams

University of Nebraska-Lincoln, nwilliams17@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/famconfacpub>

Berlin, Kristoffer S.; Parra, Gilbert R.; and Williams, Natalie A., "An Introduction to Latent Variable Mixture Modeling (Part 2): Longitudinal Latent Class Growth Analysis and Growth Mixture Models" (2013). *Faculty Publications, Department of Child, Youth, and Family Studies*. 90.
<https://digitalcommons.unl.edu/famconfacpub/90>

This Article is brought to you for free and open access by the Child, Youth, and Family Studies, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Child, Youth, and Family Studies by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

An Introduction to Latent Variable Mixture Modeling (Part 2): Longitudinal Latent Class Growth Analysis and Growth Mixture Models

Kristoffer S. Berlin,¹ Gilbert R. Parra,² and Natalie A. Williams³

1. Department of Psychology, University of Memphis

2. Department of Psychology, University of Southern Mississippi

3. Department of Child, Youth and Family Studies, University of Nebraska-Lincoln

Corresponding author — Kristoffer S. Berlin, PhD, Department of Psychology, The University of Memphis, 202 Psychology Building, Memphis, TN 38152, USA; email kristoffer.berlin@gmail.com

Abstract

Objective — Pediatric psychologists are often interested in finding patterns in heterogeneous longitudinal data. Latent Variable Mixture Modeling is an emerging statistical approach that models such heterogeneity by classifying individuals into unobserved groupings (latent classes) with similar (more homogenous) patterns. The purpose of the second of a two article set is to offer a nontechnical introduction to longitudinal latent variable mixture modeling.

Methods — 3 latent variable approaches to modeling longitudinal data are reviewed and distinguished.

Results — Step-by-step pediatric psychology examples of latent growth curve modeling, latent class growth analysis, and growth mixture modeling are provided using the Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 data file.

Conclusions — Latent variable mixture modeling is a technique that is useful to pediatric psychologists who wish to find groupings of individuals who share similar longitudinal data patterns to determine the extent to which these patterns may relate to variables of interest.

Keywords: growth mixture modeling, latent class growth analysis, latent growth curve modeling, longitudinal data analysis, structural equation modeling

Pediatric psychologists are often interested in finding patterns in heterogeneous longitudinal data. Latent Variable Mixture Modeling is an emerging statistical approach that models such heterogeneity by classifying individuals into groupings with similar patterns, called latent classes. The purpose of this second article is to offer a nontechnical overview and introduction to longitudinal mixture modeling to facilitate applications of latent variable mixture models (LVMM) within the field of pediatric psychology. In part 1 (Berlin, Williams, & Parra, 2013) we provided an overview of LVMM, highlighted the strengths of this analytic approach, and reviewed strategies for determining the optimal number of observed subgroups. Step-by-step examples were provided illustrating two prominent types of cross-sectional mixture modeling: Latent class and latent profile analyses. This companion article builds off the founda-

tional knowledge presented in part 1 and provides step-by-step examples illustrating closely related LVMMs of longitudinal data: A latent class growth analysis (LCGA) and two variants of a growth mixture model (GMM).

As described in part 1, LVMMs focus on categorical latent variables representing latent classes. LVMM is a person-centered approach that probabilistically assigns individuals into latent classes based upon similar patterns of observed cross-sectional and/or longitudinal data. LVMM groups individuals into subpopulations by inferring, based on the data, each individual's membership in latent classes. As a byproduct of mixture modeling, every individual in the data set has their own probabilities calculated for their membership in all of the latent classes estimated. Latent classes are based on these probabilities. Individuals are allowed fractional membership in all classes, reflecting the varying degrees

of certainty and precision of classification (Asparouhov & Muthén, 2007; Muthén, 2001).

In this article, we illustrate examples of longitudinal LVMM, which includes latent class growth analysis (LCGA) and growth mixture modeling. Given space limitations, longitudinal latent class models such as hidden markov models, mover-stayer models (Langeheine & van de Pol, 2002), and latent transition analyses (LTA) will not be covered; however, an introduction to and book length treatments of LTA are available for those interested (Collins & Lanza, 2009; Lanza, Flaherty, & Collins, 2003). We begin by briefly reviewing latent growth curve models (LGCMs), which serve as the foundation for both LCGAs and GMMs. We then present examples of LCGA and GMMs using the four steps recommended by Ram and Grimm (2009) with considerations from Jung and Wickrama (2007): (a) problem definition, (b) model specification, (c) model estimation, and (d) model selection and interpretation.

Latent Growth Curve Modeling

To distinguish LCGA and GMMs it is helpful to briefly review LGCMs and the parameters of this type of model. LGCM is a multivariate application of SEM that examines how individuals change on one (or more) observed outcome variable over time. In LGCM repeated measurement of observed variables are used as indicators of latent variables that represent different aspects of individual's change (see Figure 1, circles labeled intercept and slope). Most often, there are two latent variables (sometimes called random coefficients). The first is an intercept, which represents the level of the outcome when time is equal to zero and the second is a slope, which represents the rate of change in the outcome over time. In LGCM each participant has his/her own estimated intercept and slope, and these are allowed to vary across individuals. In Figure 1, this variability across individuals is estimated as the variance of the latent intercept and slope and is represented as a double-headed arrow originating from and pointing to the same variable. If the slope and intercept are believed to relate to one another, their covariance can be modeled to reflect how an individual's start value relates to his/her rate of change. Latent variables also have means, reflecting the average of all individuals' intercepts and slopes. These means are depicted in Figure 1 as the paths leading from the triangle to the intercept and slope. In addition, participants have their own deviations from these means at each time period, called residual/error variance, as well as residuals, and/or random effects, which are depicted graphically with the smaller circles with double-headed

arrows (labeled rv1–rv6). Significance tests for all these parameters are available to determine whether these estimates differ from zero and can be used to answer questions such as “*is the amount of change, on average, significantly different than zero?*” or “*is there significantly variability in individuals' rate of change?*”

The exact interpretation of the intercept and slope depend on how the researcher fixes or estimates the relations between these latent variables and their indicators. By fixing the observed variable factor loadings (e.g., the values of F1 to F6 in Figure 1), different hypothesized relations can be tested about the origin (zero point) and the rate or shape of change. This assumes that individuals were assessed at roughly the same intervals (later we will show an example when there are individually varying times of observations). Typically, the intercept factor loadings are all fixed to one (a “constant”) in tandem with fixing the first slope loading to zero (e.g., F1 in Figure 1) such that the intercept can be interpreted as the individual's estimated start value. Interpretation of the slope depends on how the remaining slope factors are specified. To model linear change, the slope loadings then are fixed to reflect the time since (or until) the zero-point. In Figure 1, children's body mass index (BMI) z-scores are measured six times over the course of eight years: Fall and spring of kindergarten, and the spring of first, third, fifth, and eighth grades. To interpret the slope as the annual linear rate of change in BMI z-scores, the factor loadings of F1–F6 would be fixed at 0, 0.5, 1.5, 3.5, 5.5, and 8.5, representing the time in years since zero. If the desired metric of time was months, these values could be fixed at 0, 6, 18, 42, and 102. Other intercept values might also be of interest, and can be modeled by altering which slope factor loading is fixed to zero. For example, by fixing the slope loading of the last time point (e.g., F6) to zero (and F1 = –8.5, F2 = –8, F3 = –7, F4 = –5, F5 = –3), the mean intercept is interpreted as the average value of the youth's BMI z-score during the spring of eighth grade, and the mean slope is interpreted as the average annual rate of change in the youth's BMI z-score.

Two of the benefits of a SEM approach to modeling change is the flexibility it offers in terms of modeling time and the available tools to evaluate these models. As a structural equation model, Figure 1 represents the researcher's hypothesis about the relation between time and the outcome, and as such, goodness-of-fit statistics and other indices can be used to establish the best way of modeling relationships and/or change over time. Figure 1 hypothesizes that changes in BMI z-scores are linear, but alternative models can be tested by adding additional latent variables, changing/freely

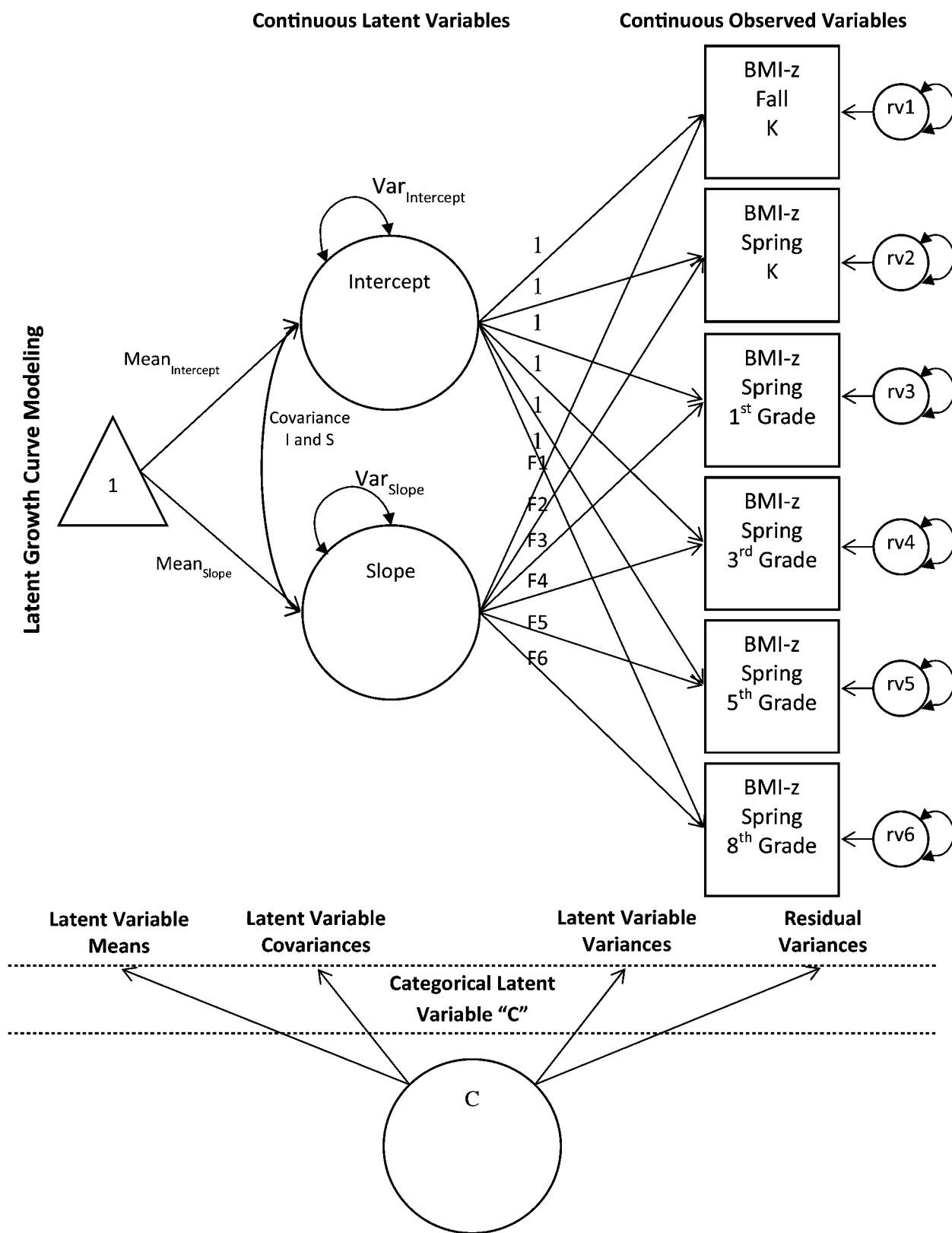


Figure 1. A Graphical Representation of a LGCM (above dashed box), and Longitudinal LVMM (above and below dashed box).

estimating the slope factor loadings, and/or imposing various constraints within the model. Alternative models include no-change (an intercept only model), polynomial change with one (quadratic function), two (cubic function), three (quartic) or more curves, piecewise (breaking growth into specific segments) and/or other complex and nonlinear models (Barker, Rancourt, & Jelalian, 2013; Grimm & Ram, 2009; Ram & Grimm, 2007). For example, if researchers believed that the amount of BMI z-score change either accelerates or decelerates with the passage of time, a third latent variable representing the quadratic slope could be added to the growth model. Factor loadings for this quadratic slope would be fixed to correspond to the square of the linear slope's loadings (i.e., 0, 0.25, 2.25, 12.25, 30.25, and 72.25). We encourage those interested in an overview of SEM specific to pediatric psychology to review Nelson, Aylward, and Steele (2008).

Two additional models warrant brief mention: Piecewise growth and latent basis models. With six or more time points, it is possible to estimate separate patterns of change for different phases of the study period (piecewise model). For example, inspection of the individual raw data and the mean level data may suggest a relatively flat pattern of change over the first three time points followed by a steep increase between the fourth and sixth time points. These models are also useful for clinical trials in which there is a period of improvement and/or symptom reduction followed by a maintenance period that has relatively little change. A benefit to the piecewise model is that it allows the researcher to estimate both trajectories simultaneously. The most flexible modeling of time is the latent basis framework (Grimm & Ram, 2009; Meredith & Tisak, 1990). This framework, along with fixing at least two time points, allows researchers to freely estimate the remaining factor loadings to reflect time that may not follow a standard mathematical shape. Often, the first loading is fixed to zero and the last is fixed to one. This allows for interpretation of the estimated factor loading as the relative percent of growth achieved by the last time period. Latent basis models may be the best choice for data that do not seem to fit typical linear, quadratic, or cubic patterns.

It is important to note that the number of time points available will influence how change over time can be modeled. With three time points, it is possible to model a linear pattern and a latent basis model; with four time points, linear and quadratic patterns can be modeled as well as a latent basis model; and with five time points, linear, quadratic, and cubic patterns can be modeled as well as a latent basis model. We encourage those interested in learning more or seeking technical details to re-

view the work of Duncan, Duncan, and Strycker (2006), Preacher, Wichman, MacCallum, & Briggs (2008), and Bollen and Curran (2005) and we refer readers to Delucia and Pitts (2006) and Singer and Willett (2003) for coverage of the multilevel approach to modeling change. We now turn our attention to modeling longitudinal data using LVMM.

Latent Class Growth Analysis and Growth Mixture Models

Both LCGA and GMMs are closely related to one another and are specific types of LGCMs. A major analytic goal of LCGA and GMM is to understand and predict individual differences (or variability) in parameters reflecting participants' change in outcomes over time. Individuals are classified into latent classes based upon similar patterns of data. The observed distribution of values may be a "mixture" of two or more subpopulations whose membership is unknown. As such, the goal of both LCGA and GMM is to probabilistically assign individuals into subpopulations by inferring each individual's membership to latent classes from the growth model data. The conceptual basis for these models is depicted graphically in Figure 1, in which the arrows from the categorical latent variable "C" to the latent means, variances, covariances, and residual variances, signifies that these parameters can vary across latent classes (to decrease clutter in the figure these arrows point to the names of the various parameters, rather than to the graphical depiction of every possible parameter).

The primary difference between LCGA and GMM is which values are allowed to vary within and across latent classes. LCGA is a special type of GMM, in which the variance of latent slope and intercept are fixed to zero within class, and allowed to vary only *across* classes. Because there is no within class variability there is no covariance between the slope and intercept, and there are far fewer parameters to estimate. With these constraints, it assumes that all individual growth trajectories within classes are homogeneous (Nagin, 1999). This approach may be particularly helpful when working with smaller sample sizes or when more complex models fail due to nonconvergence, out of range estimates, or other statistical problems, or as an initial modeling step prior to specifying a GMM model (Jung & Wickrama, 2007).

Unlike LCGA, GMM is a more flexible approach that allows researchers to determine which of the parameters depicted in Figure 1 (latent variables means, variances, covariances, and residuals, etc.) can vary both within and across classes. In the context of a LGCM model, researchers may attempt to account for vari-

ability in slopes and intercepts by adding predictors of change and baseline values to the growth model. This approach assumes that the sample is drawn from a single population that can be adequately characterized by a single set of parameters (i.e., those depicted in Figure 1; Ram & Grimm, 2009). In cases where subpopulations are *observed*, such illness type, sex, race, geographic regions, etc., this is typically handled using multiple population or multiple group approaches to LGCM (Duncan, Duncan & Strycker, 2006; Ch. 5). When heterogeneity in growth parameters may be due to the mixture of two or more unobserved subpopulations whose membership is unknown *a priori*, LGCM may not be the best approach. If these *unobserved* memberships are believed to be distinguished by different growth parameters, than LVMM may be an excellent approach to identify these individuals.

Four-Step Framework

Consistent with recommendations (Ram & Grimm, 2006), we follow a four step process to estimate LVMMs: Problem definition, model specification, model estimation, model selection and interpretation (see part 1 for an overview). The procedures unique to estimating longitudinal LVMMs are described briefly below. Example syntax is available for all models as an online supplement.

Problem Definition

When conducting LVMMs with longitudinal data, a key step in the problem definition stage is determining the best way of modeling change over time. Researchers must identify a change function (straight line [linear function], one curve [quadratic function], two curves [cubic function], or two or more separate trajectories [piecewise model]) that best represents patterns of change in the data. This single-group model will be used as the base model for the mixture analyses. Theory and prior research should be used to guide selection of the base, single-group model. In addition, inspection of individual-level raw data and mean level patterns over time can provide clues on how best to model change. If several single-group models seem plausible, a separate latent growth curve analysis should be conducted for each pattern of change. The statistical fit can then be used to evaluate each of the models and establish the best way to model change over time. To evaluate the goodness-of-fit of the single-group models, excellent models generally have the following values: CFI $\geq .95$, RMSEA $< .05$, and SRMR $< .05$ (Hu & Bentler, 1999).

Model Specification

In the model specification stage of estimating longitudinal LVMMs, hypotheses about the number of classes should be generated. Again, theory and prior research should be used to inform this decision. Researchers can also take an exploratory approach and estimate as many classes possible that yield proper solutions. In estimating longitudinal LVMMs, researchers must make initial decisions about whether they will fix the variances for the latent intercepts and latent slopes/change functions to be equal within each class. If so, a LCGM will be estimated. If the researchers want to allow within group variability for the latent intercepts and latent slopes/change functions for each class, a GMM will be estimated. In conducting a GMM, decisions about whether *means* and *variances* for latent variables (e.g., intercept and slope), *variances* and *residual variances* for observed variables, and *covariances* among latent variables will be freely estimated or fixed to be equal across classes (e.g., the parameters in Figure 1). All of these decisions can be based on theory, prior research, and/or practical considerations (model convergence, etc.).

Model Estimation

In this step, researchers select the estimation method that will be used and evaluate the statistical and conceptual fit of these models. The same strategies for cross-sectional data apply to longitudinal data (see part 1 for details).

Model Selection and Interpretation

The final model is chosen based on the various fit statistics and model considerations. The same strategies for cross-sectional data apply to longitudinal data (see part 1 for details); however, a critical step here is examining the output for nonconvergence and/or nonplausible values that might include correlations >1 and/or residual variances that are negative (which cannot be negative as the computation requires squaring values). LVMMs are known to be “finicky” and researchers using these techniques will invariably encounter such problems. Although there are many reasons for these types of errors, they are especially common if you have variables with variances that exceed 1 to 10, a group with a ceiling or floor effect, small sample/latent class sizes, and/or misspecified models. Techniques to remedy these errors have been strongly debated and include checking the tech output, increasing starts/iterations, and providing new start values. If these strategies do not resolve the problem, researchers often fix nonsignificant negative residual variances to a positive value close to zero, use a different estimator (i.e., Bayes) and/or specify a different model. These “new” models may

include fixing variances and residual variances to be equal within/across classes or freeing previously fixed parameters. In these cases it is especially important for researchers to have a clear conceptual and statistical rationale for their remedy.

Example Data

Participants

Data were obtained from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). The ECLS-K is a nationally representative sample of US children, their parents, teachers and schools. Baseline data were collected in the fall and the spring of children's kindergarten year (1998–99). Subsequently, data were collected in the fall and spring of first grade (1999–2000), the spring of third grade (2002), the spring of fifth grade (2004), and the spring of eighth grade (2007). The examples presented in this article focus exclusively on nonHispanic black participants given their elevated risk for overweight and obesity (Ogden, Carroll, Kit, & Flegal, 2012). At baseline, there were 3169 nonHispanic black children (50.2% male) and during the eighth grade assessment there were 951 (50.4% male).

Measures

Body Mass Index

Heights and weights were assessed at all six time points. These data were used to calculate BMI percentile scores that are standardized for the child's age and gender using tables provided by the Centers for Disease Control and Prevention/National Center for Health Statis-

tics (CDC, 2010). In this study, youth with BMI values ≥ 85 th to < 95 th percentile were classified as overweight, and youth with BMI values ≥ 95 th percentile were classified as obese, per existing guidelines (CDC, 2010). In addition, a standardized BMI score (BMI z-score) was calculated for each participant following guidelines established by the CDC. Biologically implausible BMI z-scores were coded as missing.

Longitudinal Latent Variable Mixture Model

Examples

Problem Definition

The first step in estimating a LVMM is identifying a single-group (nonmixture, LGCM) model that best represents change over time. Theory and prior research suggests that three to four BMI patterns will likely exist (Danner & Toland, 2013; Li, Goran, Kaur, Nollen &, Ahluwalia, 2007; Ventura, Loken and Birch 2009). Inspection of individual-level raw data (Figure 2) indicates that substantial heterogeneity that appears to be nonlinear in nature. The mean values across the six time points also suggest a nonlinear pattern of change. To ensure that we identify the model of change that best represented the six waves of data, we conducted several single-group LGCMs. These included intercept only, linear, quadratic, cubic, piecewise quadratic, and latent basis models.

Prior to conducting the LGCMs, we had to decide how to code time. For this study, data were collected at the following assessment points: Fall of kindergarten (1998–99), spring of kindergarten (1998–99), spring

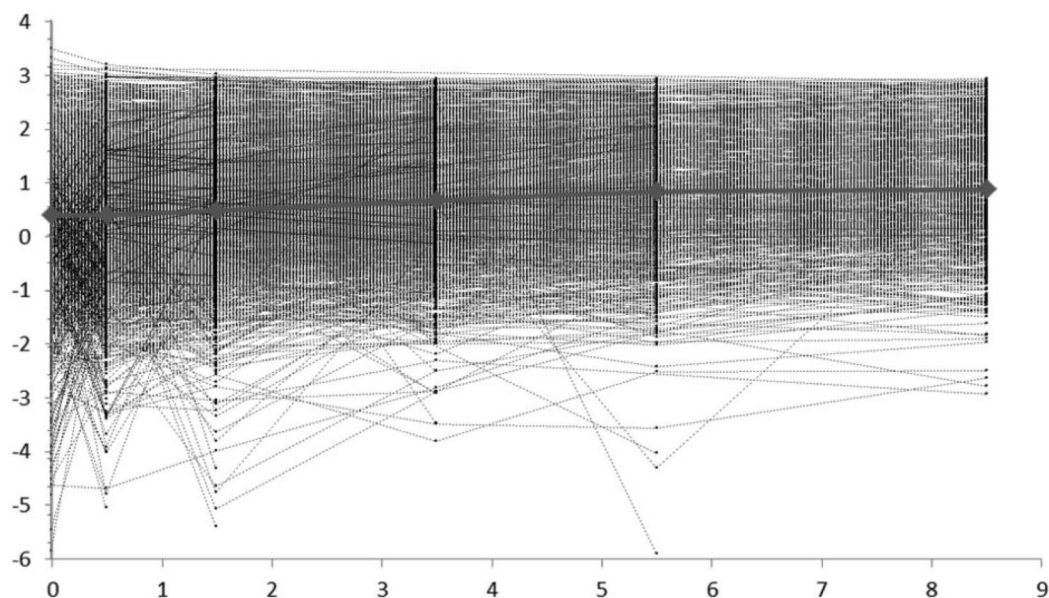


Figure 2. Observed individual African American Children's BMI z-scores kindergarten to eighth grade.

of first grade (1999-2000), spring of third grade (2002), spring of fifth grade (2004), and the spring of eighth grade (2007). We decided to code time using the following: .0, .05, .15, .35, .55, and .85. We selected this coding scheme to reflect the average time since the first measurement in the fall of kindergarten and divided each estimate by 100. In our experience, model convergence issues can result for large factor loadings for time, especially at late time points and when higher order change terms are modeled. Given that the exact date of BMI measurement was available for each participant, we also modeled time with individually-varying times of observations. This approach combines the multivariate SEM and multilevel modeling approaches to estimate random slopes and intercepts within a latent variable framework (see the TSCORES approach in Mplus; Muthén & Muthén, 1998-2012). Presently, this approach only allows for polynomial functions of time (linear, quadratic, etc.) and does not provide traditional SEM goodness of fit statistics or LVMM statistical model comparison tests discussed in part 1, such as the Lo-Mendell-Rubin test (LMR; Lo, Mendell, & Rubin, 2001) and the Bootstrap Likelihood Ratio Test (BLRT; McLachlan & Peel, 2000). All models were estimated in Mplus version 7.11 (Muthén & Muthén, 1998-2012), under missing data theory using all available data and robust (Full Information) maximum likelihood estimation. This strategy for handling missing data is an appropriate, modern method of modeling with missing data that makes use of all available data points (Little et al., 2013) and adjusts the standard errors and scales chi-square statistics to account for nonnormally distributed data. Alternative modern approaches to handling missing data were considered; however, these approaches were not chosen because they are not available within a mixture modeling framework (i.e., using auxiliary variables to predict missingness in conjunction with Full Information Maximum Likelihood) or would limit the availability of indices to help choose the optimal number of classes (e.g., model comparison likelihood ra-

tio tests discussed later are not currently available with multiple imputation techniques). To evaluate the goodness-of-fit of the single-group models, excellent models generally have the following values: CFI $\geq .95$, RMSEA $< .05$, and SRMR $< .05$. We used (robust) maximum likelihood estimation with adjusted standard errors and chi-square test statistics that are robust to nonnormality. We selected this estimator because of the skewed nature of the BMI z-scores.

As noted, we estimated the following models to identify the best single-group model that will serve as the base model for the mixture analyses: Intercept, linear, quadratic, cubic, piecewise quadratic, and latent basis (Table I). For the intercept only model, one latent factor was defined representing initial (Baseline) levels of BMI z-scores. Factor loadings for the intercepts for the six observed measures of BMI z-scores were fixed to 1. For the linear model, two latent factors were defined: One representing initial (Baseline) levels of BMI z-scores and one representing linear change in BMI z-scores (i.e., slope). Factor loadings for the intercepts for the six observed measures of BMI z-scores were fixed to 1; factor loadings for the slope factor were set to .0, .05, .15, .35, .55, .85). For the quadratic model, an additional latent variable was added to the linear model. It represented a quadratic pattern of change for BMI z-score (factor loadings fixed to .0, .0025, .0225, .1225, .3025, .7225). When estimating the quadratic model, we received a message about a negative residual variance for the eighth grade BMI z-score. Since residuals cannot be negative, we addressed this problem by fixing the residual variance for sixth time point to a small number (i.e., .00001). This solution fixed the problem and no additional warnings were generated. For the cubic model, an additional latent variable was added to the quadratic model. It represented a cubic pattern of change (factor loadings fixed to .0, .000125, .003375, .042875, .166375, .614125).

We also estimated a piecewise growth model. It included a linear slope for Time 1 through Time 3 and a quadratic function for Time 3 through Time 6 (the

Table I. Fit Statistics for Single-Group (Nonmixture) Models. ($N_s = 2543$)

Model of Change	CFI	TLI	RMSEA	SRMR	χ^2	df	χ^2/df	AIC
Intercept	0.67	0.74	0.11	.14	554.78	19	29.20	23,334.88
Linear	0.93	0.93	0.05	.04	132.55	16	8.28	22,318.98
Quadratic ^a	0.99	0.99	0.02	.02	28.08	13	2.16	22,081.74
Cubic	0.99	0.99	0.02	.01	12.88	7	1.84	22,059.32
Piecewise Quadratic	0.997	0.994	0.02	.01	11.61	7	1.66	22,057.74
Latent Basis	0.96	0.95	0.05	.06	76.95	12	6.41	22,152.07

a. The residual variance for the sixth time point was fixed to .00001. CFI = Comparative Fix Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. AIC = Akaike Information Criterion.

following Mplus code was used: `i s1| ZC1BMI@0 ZC2BMI@.05 ZC4BMI@.15 ZC5BMI@.15 ZC6BMI@.15 ZC7BMI@.15; i s2 q1| ZC1BMI@0 ZC2BMI@0 ZC4BMI@0 ZC5BMI@.2 ZC6BMI@.4 ZC7BMI@.7`). Finally, a latent basis model was estimated. For this model, factors loadings for the intercept for the six observed measures of BMI z-scores were fixed to 1. Factor loadings for the factor representing change were set to 0, *, *, *, *, 1 (asterisks indicate that a particular factor loading was freely estimated).

Based upon review of the fit statistics provided in Table I, the intercept only and linear models provided a poor fit to the data. These fit statistics also suggested that freely estimated time (latent basis), and quadratic, cubic, and piecewise models of growth all were acceptable models of nonHispanic black children's BMI z-scores from kindergarten to eighth grade. To further highlight how growth was estimated in these models, Figure 3 compares the means of the estimated BMI z-scores with the observed mean values. Visual inspection

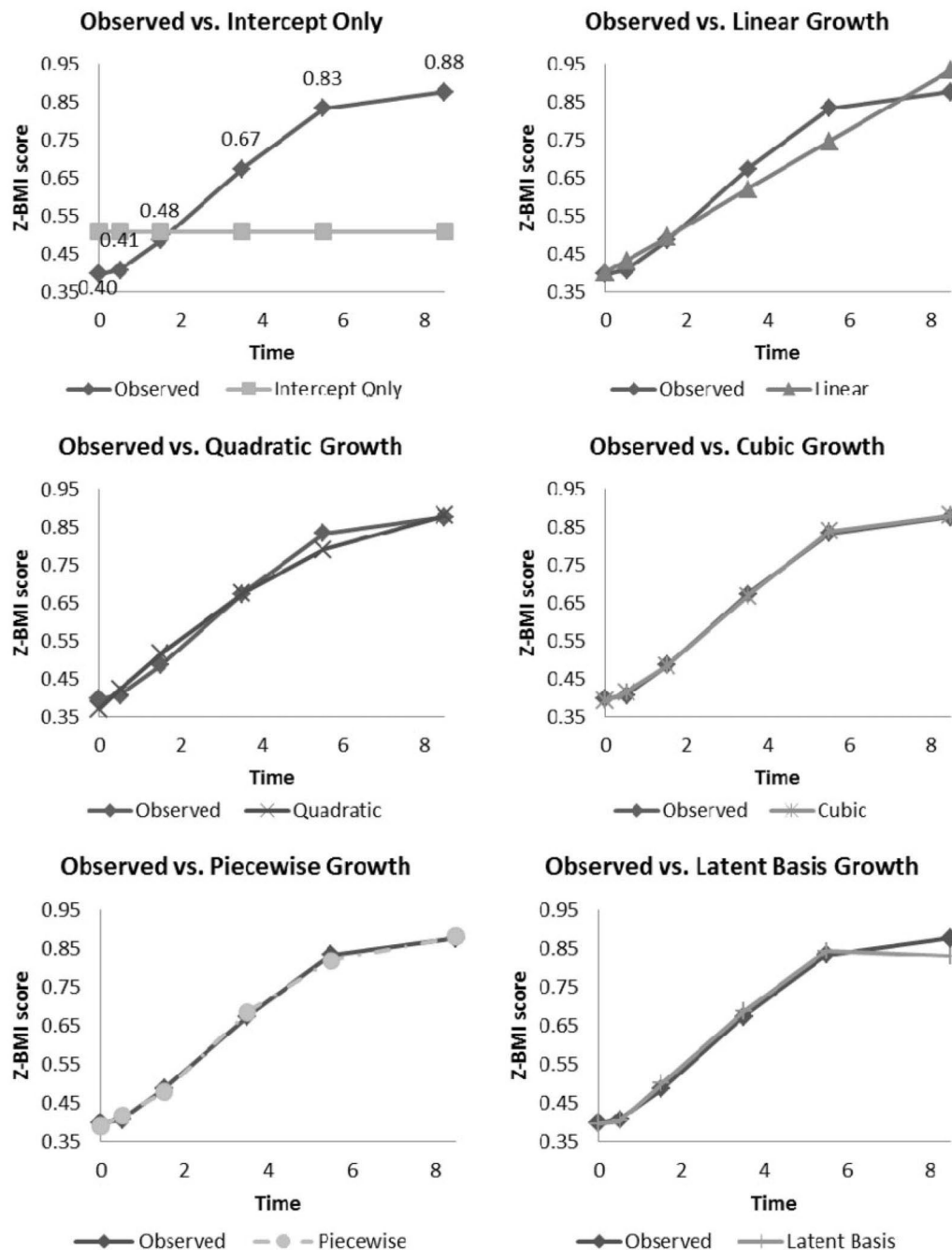


Figure 3. Observed and Estimated Model Means of African American Children's BMI z-scores.

of these graphs reveals that those models with better fit statistics tend to have a higher proportion of overlapping lines. One can see what is tested statistically, namely the discrepancy between the observed and estimated values or how well our modeled data “fit” with the actual data. Based on these findings, it was not clear which model would serve as the best base model for the growth mixture analyses. We decided to run the mixture analyses for each of the single-group models that provided an acceptable fit to the data (i.e., quadratic, cubic, piecewise quadratic, and latent basis models).

Model Specification

In this stage of estimating LVMMs, hypotheses about the number of classes are generated. We took an exploratory approach and estimated as many classes possible that yielded proper solutions. In order to provide useful examples we examined three types of LVMM: A LCGA (with the slope and intercept variances fixed to zero), and two GMMs, one with time-points that were either fixed or estimated (i.e., quadratic or latent basis), and one with individually varying times of observation. For these LVMMs, we opted for a model building approach. Specifically, for the initial GMMs that we conducted, we allowed *means* for the latent variables (e.g., intercept and slope) to vary across classes; then, we conducted analyses in which we allowed the *means* and *variances* for latent variables to vary across classes; then, we let *covariances* among the latent variable to vary across classes; finally, we allowed the variances for the six observed variables to vary across classes.

Model Estimation

As noted above, all models were estimated in Mplus version 7.11 (Muthén & Muthén, 1998–2012), under missing data theory using all available data and robust (Full Information) maximum likelihood estimation. Again, we decided to run growth mixture analyses for each of the single-group models that provided an acceptable fit to the data (i.e., quadratic, cubic, piecewise quadratic, and latent basis models). We encountered several issues when estimating the quadratic, cubic, and piecewise GMMs. One issue that emerged across each of the models was strong associations between intercepts and change functions and/or strong associations between change functions (e.g., slope and quadratic latent variables) for some classes. Our analyses often yielded messages like the following: “WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IN CLASS 2 IS NOT POSITIVE DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A LATENT VARIABLE,

A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES.” Despite considerable efforts (including increased iterations, providing start values, constraining error variances to be greater than zero, etc.), we were not able to identify a reasonable and proper fitting solution for the quadratic, cubic, or piecewise models. However, proper solutions for the latent basis models and quadratic model with individually varying times of observation were obtained and are presented below.

Proper fitting solutions were generated when using the latent basis model as the base model for the GMM and LCGA, and the quadratic model for the GMM with individually varying times of observation. As noted in the model specification section above, we ran several sets of models in which we allowed different means, variances, and covariances to vary across classes. The best solutions were obtained for a model in which the following parameters were freely estimated: (a) means and variances for latent variables representing intercept and change over time, (b) residual variances of six observed variables (BMI z-scores), and (c) covariance between intercept and change function latent variables.

Model Selection and Interpretation

We attempted to estimate two through six class solutions for all LVMMs. A five-class model did not converge on a proper solution for the latent basis GMM. Findings for latent basis LCGA, latent basis GMM, and quadratic GMM are presented in Table II and Figures 4–6. For each model (k number of classes), replication of the best Loglikelihood was verified to avoid local maxima. For models with greater than two classes, it was verified that the previous model’s loglikelihood was equal the to $k-1$ loglikelihood for the BLRT tests.

Latent Basis LCGA

Because LCGA are less complex, more clearly identify classes, and are less computationally burdensome, it is helpful to begin the model building process with a LCGA before proceeding with GMMs (Jung & Wickrama, 2007). For the present example, increasing the number of latent classes resulted in increasingly better (i.e., smaller) AICs, BICs, and SSA-BICs, without any detriment to entropy, which hovered $\sim .81$ (Table II). The model comparison tests suggested that successively adding more classes almost always resulted in a better model. Since the BIC and BLRT generally perform the best, these results suggest that the six class model should be further explored. An important caveat is that this is probably a poorly specified model, given past research finding substantial within

Table II. ICs, Entropy, Likelihood Ratio Tests for LVMMs of African American Children's z-BMIs from kindergarten to eighth grade

Measure	1 Class	2 Class	3 Class	4 Class	5 Class	6 Class
Latent Basis LCGA						
Loglikelihood	-14,902.10	-12,732.15	-11,455.31	-10,826.38	-10,432.30	-10,144.47
AIC	29,828.19	25,514.30	22,986.62	21,754.77	20,992.59	20,442.92
BIC	29,898.29	25,660.32	23,208.58	22,052.66	21,366.42	20,892.69
SSA-BIC	29,860.16	25,580.89	23,087.84	21,890.62	21,163.08	20,648.04
Entropy	N/A	0.80	0.82	0.81	0.82	0.82
LMR test	N/A	4,297.74	2,528.87	1245.63	785.04	570.08
LMR, <i>p</i> -value	N/A	<0.0001	<0.0001	0.04	0.14	0.0005
BLRT Test	N/A	4,339.90	2,553.68	1257.85	792.75	575.67
BLRT <i>p</i> -value	NA	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Latent Basis GMM						
Loglikelihood	-11061.03	-9543.72	-9333.53	9,175.88		
AIC	22,152.07	19,149.43	18,761.05	18,473.76		
BIC	22,239.68	19,330.51	19,035.58	18,830.06		
SSA-BIC	22,192.02	19,232.01	18,886.25	18,636.25		
Entropy	N/A	0.71	0.77	0.71		
LMR test	N/A	3,034.63	420.38	315.29		
LMR, <i>p</i> -value	N/A	<0.0001	=.13	<0.05		
BLRT Test	N/A	3,034.63	420.38	315.29		
BLRT <i>p</i> -value	NA	<0.0001	<0.0001	<0.0001		
Quadratic GMM*						
Loglikelihood	-11,041.044	-9,513.41	-9,316.27	-9,171.83	-9,063.08	-8,974.87
AIC	22,112.087	19,088.82	18,726.54	18,469.66	18,284.16	18,139.74
BIC	22,199.704	19,269.9	19,001.07	18,837.65	18,745.61	18,694.65
SSA-BIC	22,152.045	19,171.4	18,851.74	18,637.48	18,494.61	18,392.81
Entropy	N/A	0.717	0.618	0.635	0.602	0.59

Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Sample-Size Adjusted BIC (SSABIC), Lo-Mendell-Rubin test (LMR), BLRT, *Modeled with Individually varying time of observation, in which the LMR and BLRT are not available for these types of models.

class variability. The assumption of zero variance within classes (as modeled here) is not likely tenable and might account for each successive model (and subsequently more variability) seeming to improve the fit of the models. As such, we only briefly interpret these LCGA findings (and refer interested readers to Nagin, 1999 for more details), as the GMMs described next might be a better statistical representation of our hypothesized models of BMI z-score growth.

Latent Basis GMM

Consistent with the steps for selecting a best fitting model, we first examined the information criteria (ICs) fit statistics. As shown in Table II, IC fit statistics indicated that the four class model was the best fitting solution. We then looked at the entropy values. As shown, entropy was relatively low for all solutions, indicating that there is some inaccuracy in the classification of individuals into their most likely class. Notably, entropy values were within the acceptable range, but still quite marginal. We then examined the likelihood ratio tests.

Findings indicated that the four class solution provided the best fit to the data. Finally, we considered the substantive meaning of the solutions, and determined that the four class model made the most conceptual sense. Taken together, we selected the four class model as the best fitting solution.

Based on the growth patterns, the largest class (Class 3, 66.0%) was named "Elevated Normal Weight Tracking (ENWT). Members of this class began with an average BMI z-score of .47 (significantly different from zero, $p < .001$), that had total amount of growth across the entire time interval that was on $\sim .28$ ("slope/shape" $p < .001$). The estimated slope loading of a latent basis model reflect the proportion of the total amount of change between first (kindergarten) and last (eighth grade). Based on these factor loadings, ENWT class members had the achieved 0%, 3.5%, 31.6%, 73%, 111%, and 100% of the total z-score growth at Fall-K, and the Springs of K, first, third, fifth, and eighth grades, respectively. The second largest class (Class 2, 18.9%) named, "Rapid Increase Percentile Crossing" (RIPC) had an av-

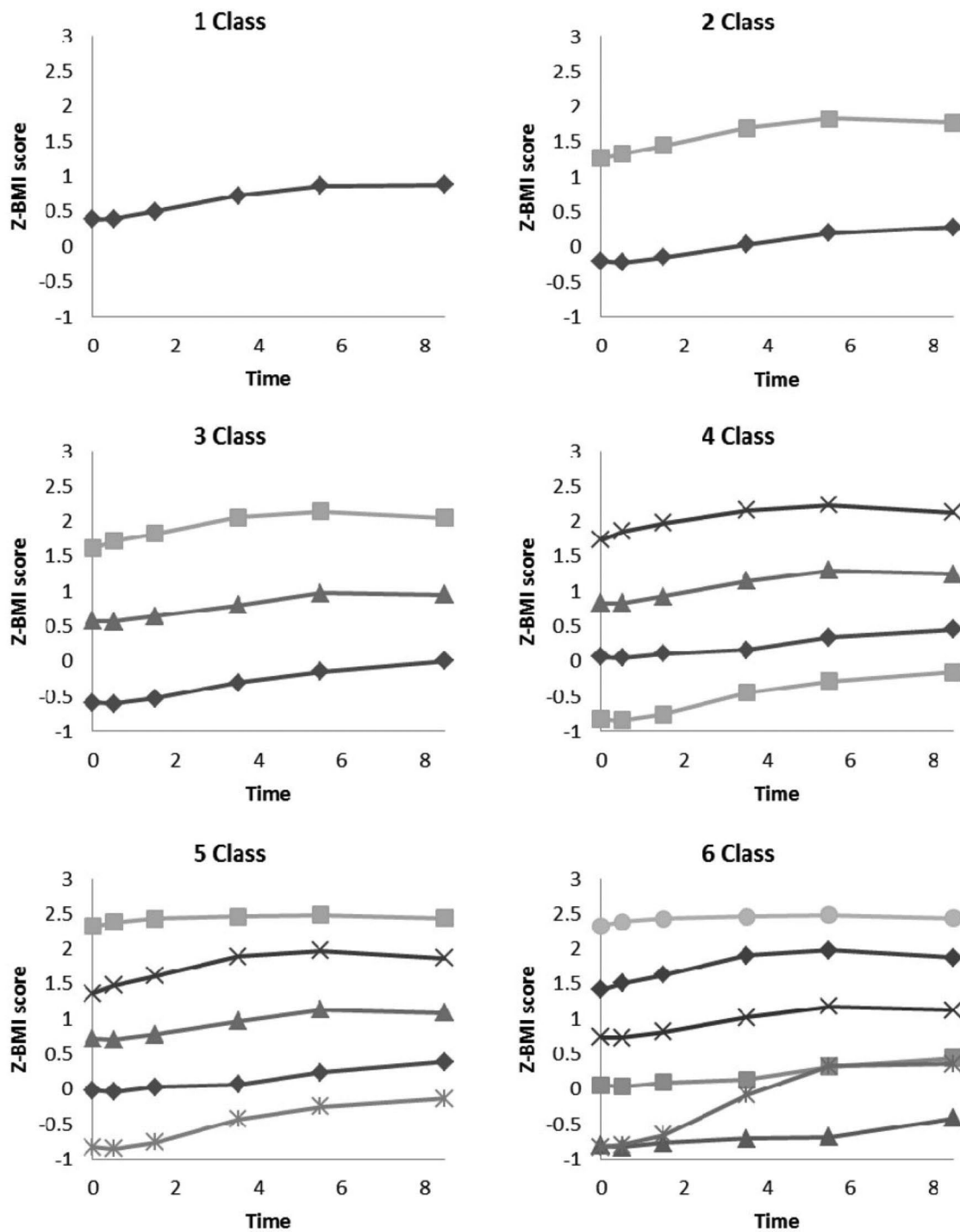


Figure 4. LCGA Models of African American Children's BMI z-scores.

average initial BMI z-score (intercept) of $-.51$ ($p < .001$), and average z-score growth total of 1.17 ($p < .001$), that was 0%, -2%, 30%, 74%, 102%, and 100% of the growth total at the six time points. The third largest class names 98th/99th Percentile Tracking (Class 4, 8.9%) had an average initial BMI z-score (intercept) of 1.98 (intercept, $p < .001$) with no significant total growth (slope = $.38$), that was at the following proportions of total growth

K through eighth grade: 0%, 39%, 73.1%, 93.7%, 98.9%, and 100%. The smallest class (Class 1, 7.2) called Mid-Childhood, Dip/Rebound, exhibited an average total growth of $.285$ (slope, $p = .05$) that started with average BMI z-scores of $.31$ ($p = .04$). Members of this class achieved the following proportion of total growth at the consecutive time points in the study: 0%, 6.1%, -211.3%, -198.8%, -48.4%, and 100%. As with previous models in

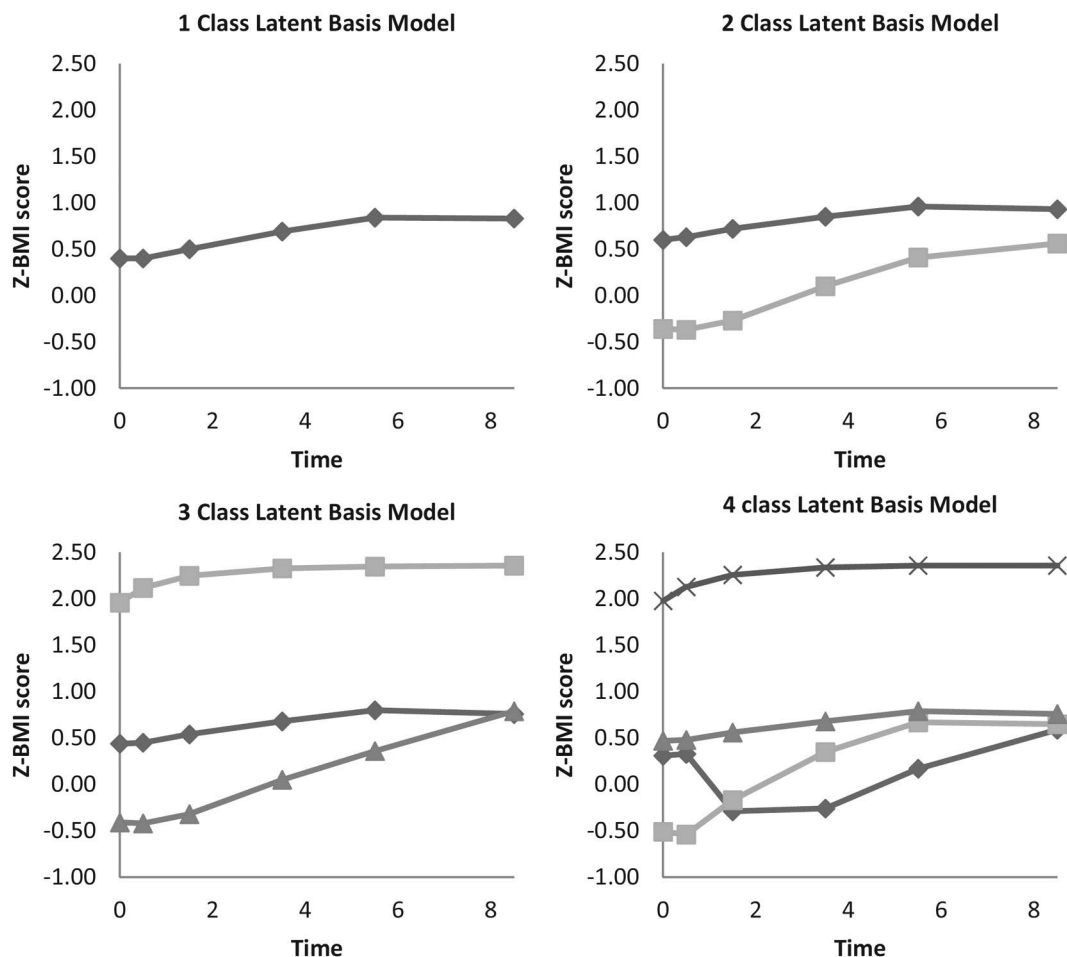


Figure 5. Latent Basis GMMs of African American Children's BMI z-scores.

part 1, we were interested in potential sociodemographic differences in the latent class. Therefore, a pseudoclass draws based multinomial logistic regression was used and revealed that the 98th/99th Percentile Tracking (Class 4) had significantly more females compared with either Class 2 ($p = .04$) or Class 3 ($p < .01$). SES was not significant predictors of latent class membership (p -values ranged from .59 to .99).

Quadratic GMM with Individually Varying Times of Observation

Quadratic GMMs were modeled specifying one through six latent classes. These GMMs freely estimated each latent classes' own latent means, intercepts, variances and covariances, and observed variable residual variances. All parameter estimates and outputs were scrutinized for out of range estimates and error messages, and when necessary additional random starts were requested to replicate the best log-likelihood value. We began by reviewing the IC. Across all ICs, adding additional classes always resulted in an improved fit (decreased val-

ues, without ever increasing). Increasing the number of classes resulted in a worsening of the classification rates, as measured by the marginal entropy ranging from .71 to .59. Inspection of "better" entropy values can be helpful when the ICs are relatively similar (Ram & Grimm, 2009). In this example, a two and four class model resulted in the best (yet still marginal) entropy values. At this point, inspection of the mean class values presented in Figure 6 can be helpful to look for overlapping class plots and/or unexpected trajectories in conjunction with our anticipation of either a three or four class model. One consideration specific to this example are proportions of youth in the most extreme weight trajectory [z -scores > 1.04 and 1.64 , which would place youth at 85th (overweight) to 95th (obese) percentile respectively]. Based on the NHANES 2000 cohort, we would anticipate the most "extreme" weight trajectory to represent 19% to 23% of the sample. Inspection of the three and four class solutions in Figure 6 reveals that both models have one such class; however, the three class model estimates that 69% of nonHispanic black youth are in the

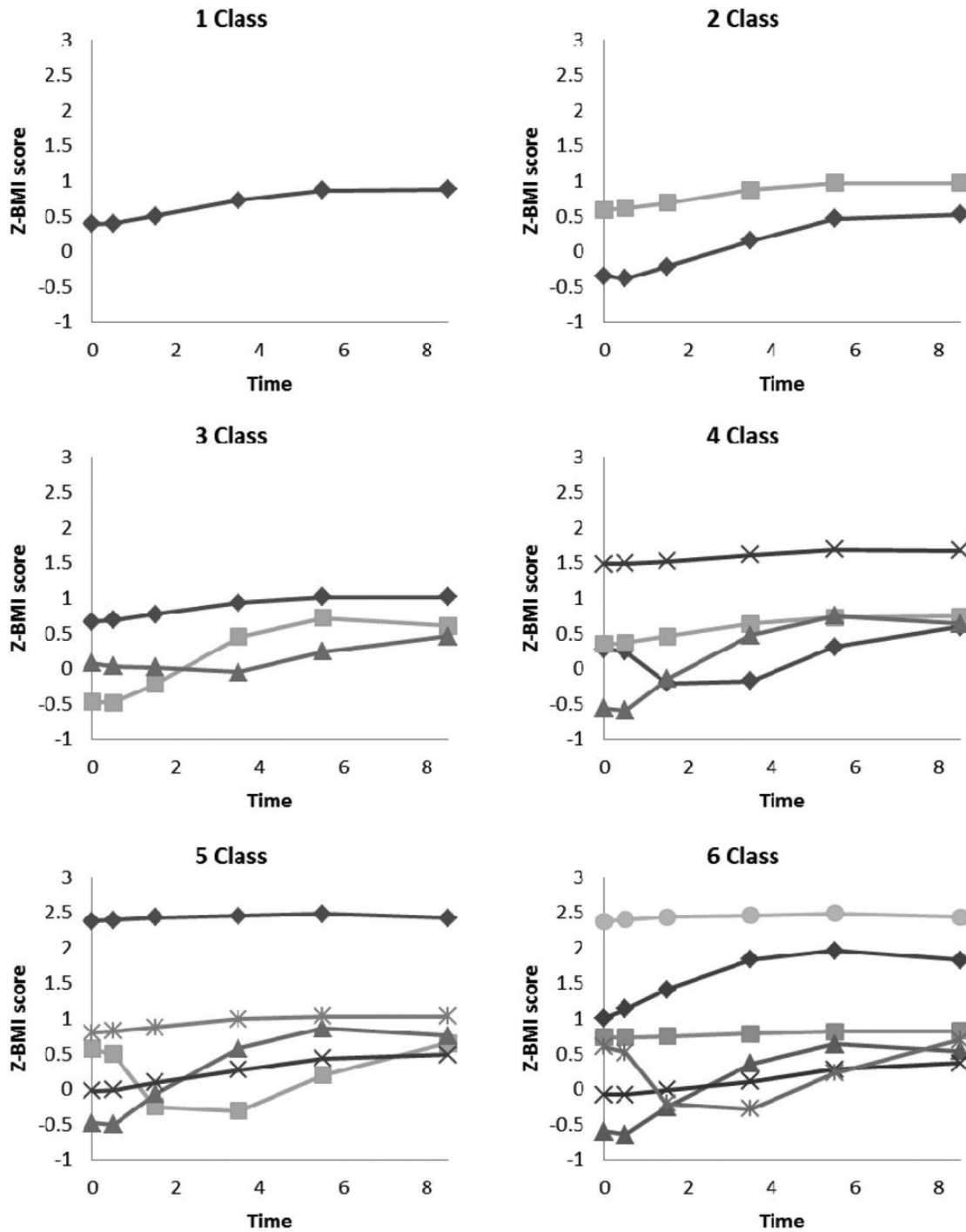


Figure 6. GMMs of African American Children's-z BMI and individually varying time of observations.

most obese trajectory, whereas, only 16% of youth are in that similar trajectory in the four class model. In light of this, the better ICs and entropy, we chose the four class solution as our final model.

Based on the growth patterns, the largest class (#2, 62.6%) was named "Elevated Normal Weight Tracking (ENWT). Members of this class began with an average BMI z-score of .34 (significantly different from

zero, $p < 0.001$), that increased (linear slope) on average .09 per year ($p < .001$) that decelerated with time (quadratic slope $-.004$, $p < .001$). Class 4, called >99th Percentile Tracking, had the second highest proportion of members (16%). Their average initial BMI z-score was 1.51 (intercept significantly different from zero, $p < .001$) with no significant linear (slope = .03, $p = .14$), or quadratic change (slope = $-.002$, $p = .34$). The third

largest class (#3, 15%) named, "Rapid Increase Percentile Crossing" (RIPC) had an average initial BMI z-score (intercept) of $-.68$ ($p < .001$), and their BMI z-score increased an average of $.04$ (linear slope) per year that decelerated with time (quadratic slope $-.03$, $p < .001$). The smallest class (#4, 7%) called Mid-Childhood Dip/Rebound started with average BMI z-scores of 1.51 (statistically different from zero, $p = .18$), that significantly decreased ($.183$, $p < .001$), reaching their lowest point in first grade then increased from third grade, reaching their highest point in eighth grade. Given this pattern of increases and decreases, neither the average linear ($.017$, $p = .14$) nor quadratic slopes ($-.002$, $p = .34$) were different from zero. A pseudoclass draws based multinomial logistic regression revealed that neither SES nor gender were significant predictors of latent class membership (p -values ranged from $.25$ to $.97$).

Comparing Latent basis LCGA, Latent Basis GMM, and Quadratic GMM

When comparing the results of the LVMM we see that LCGA trajectories have similar (parallel) slopes that are primarily distinguished as a matter of scale/severity (until the six class model), whereas both the GMMs have trajectories that are qualitatively different (more predominantly nonlinear) across classes and from the LCGA. This is due in part to the restrictions placed on LCGA, which result in a "simpler" model in which fewer parameters are estimated. This approach may be adequate given the research question that is being examined and in the context of smaller sample sizes. Also of note is the differential prediction of class membership across the types of LVMMs. Information of this kind may be helpful on deciding the final model as it may speak to validity. In our example, many sociodemographic predictors of youth's weight status have been identified (Ogden et al., 2012), and as such this information could help to decide not only the number of classes but also choosing between the variants of LCGMs and GMMs.

Conclusions

Important research questions in the field of pediatric psychology pertain to better understanding patterns of change over time. For example, How do cognitive late effects in survivors of pediatric brain tumors change over time?, How does medication adherence change following an intervention to improve behavioral functioning in adolescents with inflammatory bowel disease?, and How does family satisfaction with clinical care change following the institution of a quality improvement program aimed at reducing wait time in an outpatient pediatric sickle cell clinic? It is unlikely that all

individuals change in the same way, and as such, important subgroups of individuals many demonstrate distinct patterns of change. Identification of these meaningful subgroups and their unique patterns of change has the potential to advance current knowledge with respect to a wide range of applied research questions in pediatric psychology. Longitudinal LVMM is an analytic strategy designed to identify subgroups of individuals based on patterns of change, and thus can be a useful tool for researchers seeking to better understand the impact of health concerns on children and their families, as well as the impact of interventions designed to improve service provision, treatment adherence, and quality of life in children coping with a variety of acute and chronic health-related issues. This article was designed to provide pediatric psychologists with step-by-step instructions for carrying out two types of longitudinal LVMMs: LCGMs and GMMs. We also identified common problems that emerge when using these analytic frameworks and discussed possible solutions.

An important issue when considering whether to use LVMM is sample size, as insufficient sample sizes can be associated with convergence issues, improper solutions, and a limited ability identify small but meaningful subgroups. As discussed in part 1, sample size determination is difficult (i.e., it depends on number of parameters, missingness of data, reliability/distribution of variables, and relationship effect sizes) and rules of thumb often lead to over- or under-estimating the requisite sample size (Muthén & Muthén, 2002; Wolf et al., 2013). Monte Carlo simulations to estimate power and sample size needs (Muthén & Muthén, 2002) are likely to be a helpful approach for pediatric psychologists because they may demonstrate that small samples are sufficient under certain circumstances. Fortunately, several examples of Monte Carlo simulations designed to estimate sample size are currently available (Muthén & Muthén, 1998–2012 [example 12.3 in particular]; 2002).

In closing, we highlight valuable resources that are available for helping researchers conduct cross-sectional and longitudinal latent variable mixture modeling. These include online Mplus short courses, the Mplus Web Notes page, and the Mplus examples page—all available at <http://www.statmodel.com>. Perhaps most notable among available resources is the Mplus Discussion Board. This discussion board provides researchers an opportunity to ask the Mplus developers and their research team questions about conducting analyses using Mplus. An impressive bank of questions and responses related to LVMM is also available for review on this site, and consequently, the Mplus discussion board is often the first place we turn when faced with a chal-

lenging LVMM issue. These resources, along with this two part article, can facilitate researchers' use of LVMM to answer important questions in the field of pediatric psychology.

Supplementary Data is presented following the **References**.

Acknowledgments — The authors wish to thank Bryan T. Karazsia for serving as the associate editor for this manuscript.

References

- Asparouhov, T., & Muthén, B. O. (2007). Wald test of mean equality for potential latent class predictors in mixture modeling, [Online technical appendices]. Online: <http://www.statmodel.com/download/MeanTest1.pdf>
- Barker, D. H., Rancourt, D., & Jelalian, E. (2013). Flexible models of change: Using structural equations to match statistical and theoretical models of multiple change processes. *Journal of Pediatric Psychology*; doi: 10.1093/jpepsy/jst082
- Bollen, K. A., & Curran, P. J. (2005). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley & Sons.
- Berlin, K. S., Williams, N. A., & Parra, G. R. (2013). An introduction to latent variable mixture modeling (Part 1): Cross sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*; doi: 10.1093/jpepsy/jst084
- Collins, L. M., & Lanza, S. T. (2009). *Latent class and latent transition analysis for applications in the social, behavioral, and health sciences*. Hoboken, NJ: Wiley.
- Danner, F. D., & Toland, M. D. (2013). The interactive role of socioeconomic status, race/ethnicity, and birth weight on trajectories of body mass index growth in children and adolescents. *The Journal of Early Adolescence*, 33, 293–314; doi: 10.1177/0272431612439937
- DeLucia, C., & Pitts, S. C. (2006). Applications of individual growth curve modeling for pediatric psychology research. *Journal of Pediatric Psychology*, 31, 1002–1023.
- Duncan, T. E., Duncan, S. E., & Strycker, L. A. (2006). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in Mplus and SAS. *Structural Equation Modeling*, 16, 676–701.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jung, T., & Wickrama, K. A. S. (2007). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2, 302–317; doi: 10.1111/j.1751-004.2007.00054.x
- Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2003). Latent class and latent transition analysis. In J. A. Schinka, & W. F. Velicer, eds., *Handbook of psychology: Research methods in psychology* (pp. 663–685). Hoboken, NJ: Wiley.
- Langeheine, R., & van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenaars, & A. L. McCutcheon, eds., *Applied latent class analysis* (pp. 304–341). Cambridge, UK: Cambridge University Press.
- Li, C., Goran, M. I., Kaur, H., Nollen, N., & Ahluwalia, J. S. (2007). Developmental trajectories of overweight during childhood: Role of early life factors. *Obesity*, 15, 760–771.
- Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, W. G. (2013). On the joys of missing data. *Journal of Pediatric Psychology*; doi: 10.1093/jpepsy/jst048
- Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767–778.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Muthén, B. (2000). Latent variable mixture modeling. In G. A. Marcoulides, & R. E. Schumacker, eds., *New Developments and Techniques in Structural Equation Modeling* (pp. 1–33). Mahwah, NJ: Lawrence Erlbaum Associates.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599–620.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, 4, 139–157.

- Nelson, T. D., Aylward, B. S., & Steele, R. G. (2008). Structural equation modeling in pediatric psychology: Overview and review of applications. *Journal of Pediatric Psychology*, 33, 679–687.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage Publications.
- Ogden, C. L., Carroll, M. D., Kit, B. K., & Flegal, K. M. (2012). Prevalence of obesity and trends in body mass index among US children and adolescents, 1999–2010. *Journal of the American Medical Association*, 483–490.
- Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching method to theory. *International Journal of Behavioral Development*, 31, 303–316.
- Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying difference in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, 33, 565–576; doi: 10.1177/0165025409343765
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Ventura, A. K., Loken, E., & Birch, L. L. (2009). Developmental trajectories of girls' BMI across childhood and adolescence. *Obesity*, 17, 2067–2074.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*; doi: 10.1177/0013164413495237

SUPPLEMENT: EXAMPLE SYNTAX

```
TITLE: Example LVMM Syntax

DATA: FILE IS "zbmi.dat";

VARIABLE: !Lists the variable names
NAMES ARE
numID RACE GENDER AGE1 AGE2 AGE4 AGE5 AGE6 AGE7 WKSESL W8SESL
alc a2c a4c a5c a6c a7c !Age alc centered to reflect time since average of zero in Kindergarten;

USEOBSERVATION ARE (RACE EQ 2); !USED to select only African American Children
USEVARIABLES ARE
  ZC1BMI ZC2BMI ZC4BMI ZC5BMI ZC6BMI ZC7BMI;
IDVARIABLE = numID;
MISSING ARE ALL (-99);

CLASSES = c (3); ! Used with "type = mixture;" Number in parentheses is the number of requested classes

ANALYSIS:
  ESTIMATOR IS MLR;          ! Specifies Robust ML, used for all models, other estimators are available
  !TYPE = MIXTURE;          ! Used for LCGA and GMM
  !TYPE = RANDOM MIXTURE;    ! Used for GMM with individually varying times of observation

MODEL:

!LATENT GROWTH CURVE MODELS:

! i | ZC1BMI@0 ZC2BMI@.05 ZC4BMI@.15 ZC5BMI@.35 ZC6BMI@.55 ZC7BMI@.85; ! Intercept Only
! i s | ZC1BMI@0 ZC2BMI@.05 ZC4BMI@.15 ZC5BMI@.35 ZC6BMI@.55 ZC7BMI@.85; ! Linear
! i s q | ZC1BMI@0 ZC2BMI@.05 ZC4BMI@.15 ZC5BMI@.35 ZC6BMI@.55 ZC7BMI@.85; !Quadratic
! ZC7BMI@0000.1; ! FIXES the residual variance to 00001);
! i s q c | ZC1BMI@0 ZC2BMI@.05 ZC4BMI@.15 ZC5BMI@.35 ZC6BMI@.55 ZC7BMI@.85; !Cubic
! i s1 | ZC1BMI@0 ZC2BMI@.05 ZC4BMI@.15 ZC5BMI@.15 ZC6BMI@.15 ZC7BMI@.15; !Piecewise part 1 (needs part 1 & 2)
! i s2 q2 | ZC1BMI@0 ZC2BMI@0 ZC4BMI@0 ZC5BMI@.2 ZC6BMI@.4 ZC7BMI@.7; !Piecewise part 2 (needs part 1 & 2)
! i s | ZC1BMI@0 ZC2BMI* ZC4BMI* ZC5BMI* ZC6BMI* ZC7BMI@1; ! Latent Basis

!LATENT CLASS GROWTH ANALYSIS MODELS (used with type = mixture;)
!%OVERALL%
! i s | ZC1BMI@0 ZC2BMI*0.05 ZC4BMI*0.15 ZC5BMI*.035 ZC6BMI*.55 ZC7BMI@1; ! Latent Basis
! i@0; s@0; !Fixes the variance of the intercept and slope to zero
!%C#1% ! Model specific to class 1
! i s | ZC1BMI@0 ZC2BMI*0.05 ZC4BMI*0.15 ZC5BMI*.035 ZC6BMI*.55 ZC7BMI@1; ! !frees this classes' factor
loading to vary
! ZC1BMI* ZC2BMI* ZC4BMI* ZC5BMI* ZC6BMI* ZC7BMI*; !frees the residual variance to differ across classes
! [i s]; !frees this classes' latent variable means to vary

!GMM WITH INDIVIDUALLY VARYING TIMES OF OBSERVATION (used with type = random mixture;)

!%OVERALL% !Specifies the overall model;
!i s q | zb1 zb2 zb4 zb5 zb6 zb7 AT alc a2c a4c a5c a6c a7c; ! Quadratic Growth Tscores, Overall model
!i s q with i s q; i; s; q; zb1; zb2; zb4; zb5; zb6; zb7;

!%C#1% !SPECIFIC estimates for the class number 1, repeated for each class, by changing C#2%, %C#3%, etc.
! [i s q]; !Freely Estimated latent means
! i s q with i s q; !Freely Estimated covariances
! i; s; q; !Freely Estimated variances
! zb1; zb2; zb4; zb5; zb6; zb7; !Freely Estimated residual variances

!GROWTH MIXTURE MODEL: Latent Basis

!%OVERALL%
! i s | ZC1BMI@0 ZC2BMI*.05 ZC4BMI*.15 ZC5BMI*.35 ZC6BMI*.55 ZC7BMI@1;
! [i s]; !Freely Estimated latent means
! i with i s; !Freely Estimated covariances
! i; s; !Freely Estimated variances
! zb1; zb2; zb4; zb5; zb6; zb7; !Freely Estimated residual variances
! %C#1% !parameters to be estimated differently for designated class
! i s | ZC1BMI@0 ZC2BMI*.05 ZC4BMI*.15 ZC5BMI*.35 ZC6BMI*.55 ZC7BMI@1;
! [i s]; !Freely Estimated latent means
! i with i s; !Freely Estimated covariances
! i; s; !Freely Estimated variances
! zb1; zb2; zb4; zb5; zb6; zb7; !Freely Estimated residual variances

PLOT: SERIES = ZC1BMI(0) ZC2BMI(.05) ZC4BMI(.15) ZC5BMI(.35) ZC6BMI(.55) ZC7BMI(.85);
TYPE = PLOT3;
```